Scale-Free Overlay Topologies with Hard Cutoffs for Unstructured Peer-to-Peer Networks

Hasan Guclu
Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545. guclu@lanl.gov

Murat Yuksel
Computer Science and Engineering Department
University of Nevada - Reno
Reno, NV 89557. yuksem@cse.unr.edu

Abstract

In unstructured peer-to-peer (P2P) networks, the overlay topology (or connectivity graph) among peers is a crucial component in addition to the peer/data organization and search. Topological characteristics have profound impact on the efficiency of search on such unstructured P2P networks as well as other networks. A key limitation of scale-free (power-law) topologies is the high load (i.e. high degree) on very few number of hub nodes. In a typical unstructured P2P network, peers are not willing to maintain high degrees/loads as they may not want to store large number of entries for construction of the overlay topology. So, to achieve fairness and practicality among all peers, hard cutoffs on the number of entries are imposed by the individual peers, which limits scale-freeness of the overall topology. Thus, it is expected that efficiency of the flooding search reduces as the size of the hard cutoff does. We investigate construction of scale-free topologies with hard cutoffs and effect of these hard cutoffs on the search efficiency.

1. Introduction

In decentralized P2P networks, the overlay topology (or connectivity graph) among peers is a crucial component in addition to the peer/data organization and search. Topological characteristics have profound impact on the efficiency of search on P2P networks as well as other networks. It has been well-known that search on small-world topologies can be as efficient as $O(\ln N)$ [17], and this phenomenon has recently been studied on P2P networks [20, 15, 16]. The best search efficiency in realistic networks can be achieved when the topology is scale-free (power-law), which offers search efficiencies like $O(\ln \ln N)$. However, generation and maintenance of such scale-free topologies are hard to realize in a distributed and potentially uncooperative environments as in the P2P networks. Another key limitation of scale-free topologies is the high load (i.e. high degree) on very few number of hub nodes. In a typical unstructured P2P network, peers are not willing to maintain high degrees/loads as they may not want to store large number of entries for construction of the overlay topology. So, to achieve fairness and practicality among all peers, *hard* cutoffs on the number of entries are imposed by the individual peers. These hard cutoffs might limit *scale-freeness* of the overall topology, by which we mean having a network with a power-law degree distribution from which an exponent can be obtained properly. Thus, it is expected that the search efficiency reduces as the size of the hard cutoff does.

Although scale-free topologies are superior in search efficiency, their super-hub-based structure makes them vulnerable to threats and impractical due to unfair assignment of network load on a very small subset of all nodes. As peers in a P2P network are typically not fully cooperative, protocols cannot rely on methods working with full cooperation of peers. For example, peers may not want to store large number of entries for construction of the overlay topology, i.e. connectivity graph. Even though characteristics of the overlay topology is crucial in determining the efficiency of the network, peers typically do not want to take the burden of storing excessive amount of control information for others in the network. Effect of this on the overlay topology maintenance is that peers impose *hard cutoffs* on the amount of control information to be stored. Since P2P overlay topology generation and maintenance are very important for realizing a scalable unstructured P2P network, the main focus of this paper is *to investigate the effect of the hard cutoffs on the overall search efficiency*.

A key issue is the construction of scale-free overlay topologies without global information. There are several techniques to generate a scale-free topology, which rely on *global* information about the current network when a new node joins. Such global methods are not practical in P2P networks, and *local* heuristics in generating such scale-free overlay topologies with hard cutoffs is the key issue, which we investigate in this paper. In other words, each peer has

27th International Conference on Distributed Computing Systems (ICDCS'07)
0-7695-2837-3/07 $20.00 © 2007 IEEE

to figure out the optimal way of joining the P2P overlay by only using the locally available (i.e. immediate/close neighbors) information, and also causing a minimal inefficiency to the search mechanisms to be run on the network.

This paper touches an uncovered set of research problems relating to tradeoffs between maximum number of links a peer can (or is willing to) store and the efficiency of search on an overlay topology composed of such peers. We defined the maximum number of links to be stored by peers as the *hard cutoff* for the degree of a peer in the network as compared to *natural cutoff* which occurs due to finite-size effects. The primary focus of this paper is to (i) investigate construction of scale-free topologies with hard cutoffs (i.e. there are not any major hubs) in a distributed manner without requiring global topology information at the time when nodes join and (ii) to investigate the effect of these hard cutoffs on the search efficiency.

The rest of the paper is organized as follows: First, we survey previous work on P2P networks in Section 2. In Section 3, we survey the previous work on scale-free topology generation. We introduce our practical topology generation methodologies, Hop-and-Attempt Preferential Attachment (HAPA) and Discover-and-Attempt Preferential Attachment (DAPA), in Section 4. In Section 5, we present our simulations of three different search algorithms (i.e., Flooding (FL), Normalized Flooding (NF), and Random Walk (RW)) on topologies generated by the introduced models. We conclude by summarizing the work and future directions in Section 6.

2. Related Work

Our work is related to peer-to-peer (P2P) network protocol designs and topological analysis of complex networks. Previous work on P2P network protocols can be classified into *centralized* and *decentralized* ones. As centralized P2P protocols (e.g. Napster [3]) proved to be unscalable, the majority of the P2P research has focused on decentralized schemes. The decentralized P2P schemes can be further classified into sub-categories: *structured*, *unstructured*, and *hybrid*. In the structured P2P networks, data/file content of peers is organized based on a keying mechanism that can work in a distributed manner, e.g. Distributed Hash Tables (DHTs) [22]. In contrast to the structured schemes, unstructured P2P networks do not include a strict organization of peers or their content. Since there is no particular keying or organization of the content, the search techniques are typically based on flooding. Thus, the searches may take very long time for rare items. To balance the tradeoffs between structured and unstructured schemes, hybrid approaches (e.g. [23, 26]) have attempted attaining a middle-ground between the costly maintenance of global peer/data keying of structured schemes and the high cost searches of unstructured schemes.

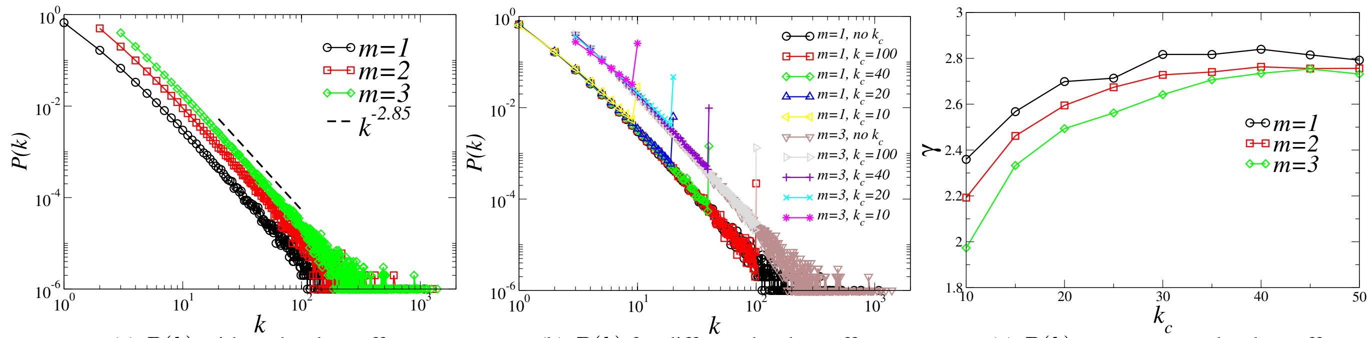
Since our work is more applicable to unstructured P2P networks, we focus our survey in this section to that category of the previous work. The main focus of the research on unstructured P2P networks has been the tradeoff between state complexity of peers (i.e. number of records needed to be stored at each peer) and flooding-based search efficiency. The minimal state each peer has to maintain is the *list of neighbor peers*, which construct the overlay topology. Optionally, peers can maintain *forwarding table*s (also referred as routing tables in the literature) for data items in addition to the list of neighbor peers. Thus, we can classify unstructured P2P networks into two based on the type(s) of state peers maintain: (i) *per-data* unstructured P2P networks (i.e. peers maintain both the list of neighbor peers and the per-data forwarding table), and (ii) *non-per-data* unstructured P2P networks (i.e. peers maintain only the list of neighbor peers).

In non-per-data schemes, search is performed by means of flooding query packets. Search performance over such P2P networks has been studied in various contexts, which includes pure random walks [13], probabilistic flooding techniques [18, 14], and systematic filtering techniques [25].

Per-data schemes (e.g. Freenet [2]) can achieve better search performances than non-per-data schemes, though they impose additional storage requirements to peers. By making the peers maintain a number of <key,pointer> entries peers direct the search queries to more appropriate neighbors, where "key" is an identifier for the data item being searched and the "pointer" is the next-best neighbor to reach that data item. This capability allows peers to leverage associativity characteristics of search queries [10].

3. Scale-Free Network Topologies

Recent research shows that many natural and artificial systems such as the Internet, World Wide Web have power-law (i.e. scale-free) degree (connectivity) distributions, i.e., $P(k) \sim k^{-\gamma}$. Barabási and Albert [6] proposed a mechanism known as *preferential attachment* (PA or rich get richer) which generates a scale-free network for which the degree distribution exponent γ=3. In this study, we use a simple version of the PA model. The model evolves by one node at a time and this new node is connected to m (number of stubs) different existing nodes with probability proportional to their degrees, i.e., $P_i = k_i / \sum_j k_j$ where k_i is the degree of the node i. The average degree per node in the resulting network is $2m$. Fig. 1(a) shows the degree distributions of scale-free networks generated by the PA model with different m values. The power-law fits to the distributions have exponents between $(-2.9, -2.8)$. The dashed line is a power-law function with exponent $\gamma = 2.85$. Sim-



(a) $P(k)$ without hard cutoff. (b) $P(k)$ for different hard cutoffs. (c) $P(k)$ exponents vs. hard cutoffs.

Figure 1. Degree distribution $P(k)$ for various networks generated by the PA model. The number of nodes is $N{=}10^5$ and for every data point 10 different realizations of the network have been used.

Table 1. Scale-freeness vs. network diameter

Diameter d	Exponent γ	Number of stubs m
$O(\ln\ln N)$	(2,3)	≥ 1
$O(\ln N/\ln\ln N)$	3	≥ 2
$O(\ln N)$	3	1
$O(\ln N)$	> 3	≥ 1

ulations show that the exponent $\gamma = 3$ is attainable only for very large networks. The special case of the PA procedure is when the number of stubs is one (i.e. m=1) in which a scale-free tree without clustering (loops) is generated.

Scale-free networks also have *small-world* [24] properties. In small-world networks the diameter, or the mean hop distance between the nodes scales with the system size (or the number of network nodes) N logarithmically, i.e., $d \sim \ln N$. The scale-free networks with $2 < \gamma < 3$ have a much smaller diameter and can be named *ultra-small* networks [11], behaving as $d \sim \ln\ln N$. When $\gamma = 3$ and $m \geq 2$, d behaves as in $d \sim \ln N/\ln\ln N$. However, when $m = 1$ for $\gamma = 3$ the Barabási-Albert model turns into a tree and $d \sim \ln N$ is obtained. Also when $\gamma > 3$, the diameter still behaves logarithmically $d \sim \ln N$. These relationships are summarized in Table 1. Since the speed/efficiency of search algorithms strongly depend on the average shortest path, scale-free networks have much better performance in search than other random networks.

3.1. The Cutoff

One of the important characteristics of scale-free networks is the natural cutoff degree (or maximum degree) due to the finite-size effects. In [5] it was defined as the value of the degree, for which the expected number of nodes is 1, i.e., $N\text{Prob}(k = k_{nc}){\sim}1$. For a scale-free network, when one substitutes $P(k){\sim}k^{-\gamma}$ into equation above, $k_{nc}(N) \sim N^{1/\gamma}$ is obtained. This definition of natural cutoff degree used in many P2P network studies lacks some mathematical rigor since it considers the probability of a single point in a probability distribution, which is not completely well-defined in the continuous k limit for large N [8]. A more physical definition of cutoff was given in [12] as the value of the degree above which one expects to find at most one vertex. Namely,

$$N \int_{k_{nc}}^{\infty} P(k)dk \sim 1. \tag{1}$$

By using the exact form of probability distribution (i.e., $P(k) = (\gamma - 1)m^{\gamma-1}/k^{\gamma}$), one obtains

$$k_{nc}(N) \sim mN^{1/(\gamma-1)}. \tag{2}$$

For the scale-free networks generated by PA model (γ=3) the natural cutoff becomes

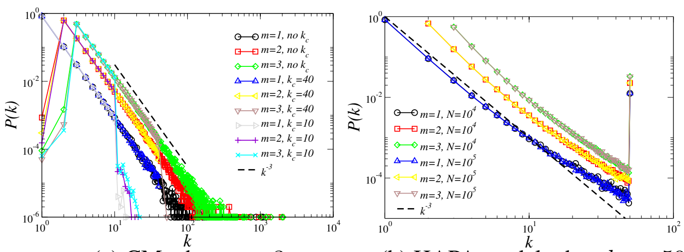
$$k_{nc}(N) \sim m\sqrt{N}. \tag{3}$$

3.2. Preferential Attachment (PA) with Hard Cutoffs

The natural cutoff may not be always attainable for most of the scale-free networks due to technical reasons. One main reason is that the network might have limitations on the number of links the nodes can have. This is especially important for P2P networks in which nodes can not possibly connect many other nodes. This requires putting an artificial or *hard* cutoff k_c to the number of links one node might have.

In order to see the effect of hard cutoff in PA, we simply did not allow nodes to have links more than a fixed hard cutoff value during the attachment process. This modified method generates a scale-free network in which there are many nodes with degree fixed to hard cutoff instead of a few very high-degree hubs and the degree distribution still decays in a power law fashion. The degree distributions of scale-free networks generated by PA with different hard cutoff values are shown in Fig. 1(b). As can be seen in the figure they are slightly different than PA without a cutoff

27th International Conference on Distributed Computing Systems (ICDCS'07)
0-7695-2837-3/07 $20.00 © 2007 IEEE



(a) CM when $\gamma = 3$ (b) HAPA model when $k_c = 50$

Figure 2. Degree distribution for CM and HAPA model.

in terms of exponent except that they have an accumulation of nodes with degree equal to hard cutoff. PA model, in its original form, has a constant degree distribution exponent γ=3 for very large networks. However, when a hard cutoff is introduced it is observed that the absolute value of degree distribution exponent decreases as in Fig. 1(c).

3.3. Configuration Model (CM)

Given that the PA model yields lower degree distribution exponents as the hard cutoff reduces, we were motivated to work on generation of power-law networks with different exponents. In other words, instead of dealing with hard cutoffs, it is possible to achieve a natural cutoff value less than the targeted hard cutoff. In this manner, the peaks at the hard cutoff value of Fig. 1(b) can be prevented and a smooth power-law distribution of degrees can be obtained. Here, we use the configuration model (CM) with a predefined degree distribution to generate a static scale-free network [21].

CM [7, 21] was introduced as an algorithm to generate uncorrelated random networks with a given degree distribution. In CM, the vertices of the graph are assigned a fixed sequence of degrees $\{k_i\}_{i=0}^{N}, m \leq k_i \leq k_c$, where typically k_c=N, chosen at random from the desired degree distribution $P(k)$, and with the additional constraint that the $\sum_i k_i$ must be even. Then, pairs of nodes are chosen randomly and connected by undirected edges. This model generates a network with the expected degree distribution and no degree correlations; however, it allows self-loops and multiple-connections when it is used as described above. It was proved in [8] that the number of multiple connections when the maximum degree is fixed to the system size, i.e., k_c=N, scales with the system size N as $N^{3-\gamma} \ln N$. Since we work with hard cutoff values typically less than the natural cutoff the number of multiple links is much less than the original CM for which k_c=N [9]. After this procedure we simply delete the multiple connections and self-loops from the network which gives a very marginal error in the degree distribution exponent. Deleting this discrepancies also causes some very negligible number of nodes in the network to have degrees less than the fixed minimum degree (m) value, even zero, as seen in Fig. 2(a). One another characteristic of the CM is that the network is not a connected network when m=1, i.e., it has disconnected clusters (or components). For m>1, the network is almost surely connected having one giant component including all the nodes.

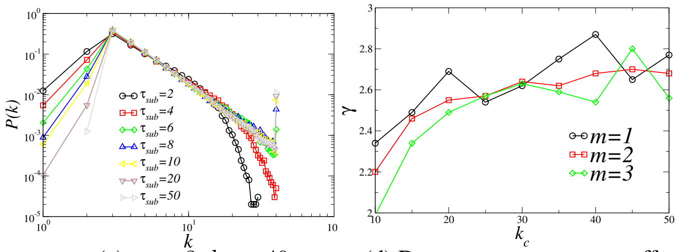
4. Local Heuristics for Scale-Free Overlay Topology Construction

In the PA model as outlined in the previous section, the new node has to make random attempts to connect to the existing nodes. To implement this in a P2P (or any distributed) networks, the new node has to have information about the global topology (e.g., current total degree in the network and the degree of the node it attempts to connect), which might be very hard to maintain in reality. Such global topology information is needed in the CM as well. Thus, in order for a topology construction mechanism to be practical in P2P networks, it must allow joining of new nodes by just using locally available information. Of course, the cost of using only local information is expected to be loss of scale-freeness (or any other desired characteristics) of the whole overlay topology, which will result in loss of search efficiency in return. In this section, we present two practical methods using local heuristics not necessarily using global information about the latest topology: HAPA and DAPA.

4.1. Hop-and-Attempt PA (HAPA)

In this method, the new node randomly selects an existing node and attempts to connect. Then it randomly selects a node which is a neighbor of the previously selected node and attempts to connect. Thus, the new node hops between the neighboring nodes and attempts to connect by using the existing links in the network until it fills all its stubs, i.e., the number of links it has reaches m.

This hopping process gives a better chance to the new node to find the high-degree hubs in the network than the PA does since the hubs in scale-free networks are only a couple of hops away from the low-degree nodes and it is less likely to find hubs by random node selection. So, some nodes in the network (probably they are the initial nodes and their number is $m + 1$ due to network generation algorithm) become dominant and attract almost all the nodes to themselves, thus deserve the name *super hubs*. The super hubs have degrees on the order of network size. It is easily seen that this procedure makes the topology of the system a star-like topology if the network is not limited by a cutoff. Naturally, without a hard cutoff the degree distribution is not a power-law and the average shortest path/diameter is very small with respect to scale-free networks generated by PA. As shown in Fig. 2(b), when a hard cutoff is introduced the degree distribution gets closer to a power law having an



(a) $m = 3$, $k_c = 40$ (d) Degree exponent vs. cutoff

Figure 3. Degree distribution of DAPA model.

exponent $\gamma = 3$ but with possibly exponential factors making a degree exponent very hard.

4.2. Discover-and-Attempt PA (DAPA)

DAPA model imitates the method for finding peers in Gnutella-like [1] unstructured P2P networks. First, we assume that we have a network called *substrate* network with a predefined and preconstructed topology at hand. We used geometric random networks as the substrate network. Then, we construct an overlay network on this substrate network by using the PA method among the set of nodes visible/reachable to a specific node (the horizon of the node) in a number of steps, which we call *local time-to-live* (or local TTL) and represent with τ_{sub}.

In this model, initially, a few nodes are randomly selected from the substrate network and added to the previously empty overlay network, G_O, then these nodes are connected to each other in G_O. At each step one random node is chosen in the substrate network and let it send a query to its neighborhood reachable in τ_{sub} hops to get a list of peers in its horizon. Then by using the rules of preferential attachment the new node connects to m peers with probability proportional to their degrees divided by the total degrees of the peers in its horizon. If the number of peers in the horizon is less than m, then the new node connects to peers it can find. The nodes which can find at least one peer in their horizon is added to the G_O and becomes a peer. A peer belonging to G_O can not be selected again to look for new peers. This process is continued until the number of peers in G_O reaches the desired number N_O.

The degree distribution of the network generated by DAPA model exhibits some interesting characteristics. For small values of τ_{sub}, the nodes are shortsighted, i.e., they cannot see enough peers in their short horizon causing the degree distribution to be an exponential. For high enough τ_{sub} values, the degree distribution changes into a power-law. Thus, one can go from an exponential to a scale-free network by playing with the measure of locality (τ_{sub}), as can be seen in Fig. 3. As the hard cutoff gets smaller the difference between the degree distributions becomes invisible. For higher values of m (i.e., $m > 1$), it is possible to find peers with degree less than m as in Fig. 3(a), since some nodes cannot find enough peers in their horizon to fill all their stubs. The degree distribution exponent has a similar behavior to PA as we change the hard cutoff value, i.e., as the cutoff decreases the exponent increases [see Fig. 3(b)]. The data in Fig. 3(b) is very noisy and the data points contain quite large error bars because they are obtained from very scattered degree distribution tails.

When a peer is to join the current overlay topology, the PA and CM do need global information about the current topology whilst HAPA and DAPA methods use local information partially or fully, respectively. Therefore, HAPA and DAPA methods are more practical in the context of unstructured P2P networks.

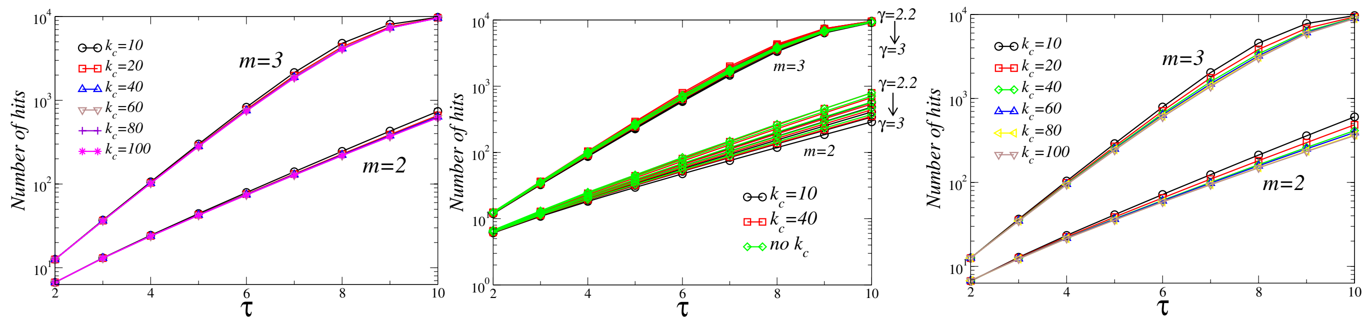
5. Simulations

We study a number of search algorithms that can be used to search items in P2P networks utilizing the power-law (the presence of hubs) degree distribution in sample networks generated by our topology construction algorithms. We consider three different search algorithms: *flooding* (FL), *normalized flooding* (NF), and *random walk* (RW).

5.1. Search Algorithms

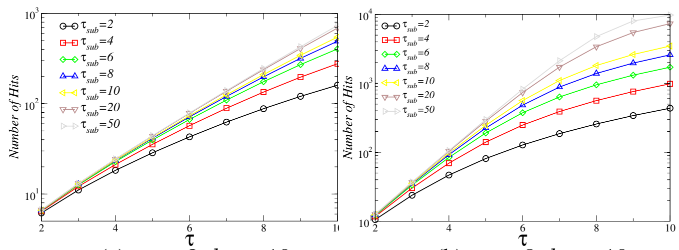
Flooding (FL): FL is the most common search algorithm in unstructured P2P networks. In search by FL, the source node s, sends a message to all its nearest neighbors. If the neighbors do not have the requested item, they send on to their nearest neighbors excluding the source node. This process is repeated a certain number of times, which is usually called as *time-to-live (TTL)* and we represent it with τ in this paper. After a message is forwarded an amount of time equal to τ, it is discarded. Independent floods by the nodes make the FL algorithm parallel. On the other hand, in this algorithm a large number of messages is created since the destination node cannot stop the search. This corresponds to a complete sweep of all the nodes within a τ hop distance from the source. The delivery time in search by FL is measured of intermediate links traversed, and is equal to the shortest path length. Since the average shortest path for small-world networks, including scale-free ones generated by the PA model, is proportional to the logarithm of system size N or even slower, the average delivery time (T_N) is logarithmic as well, i.e., $T_N = \log(N)$.

Normalized Flooding (NF): In search by FL, when large degree nodes (hubs) are reached, the number of neighbors for the next step in FL increases dramatically leading to a poor granularity. This also causes a lot of shared edges reducing the performance in terms of number of messages per distinct number of discovered nodes. To overcome this problem, search by NF algorithm was introduced in [14]. In NF, the minimum degree in the network k_{min} is an important factor. NF search algorithm proceeds as follows:



(a) PA model (b) CM (c) HAPA model

Figure 4. Normalized Flooding results for PA, CM, and HAPA models when $m = 2$ and $m = 3$.



(a) $m = 2$, $k_c = 10$ (b) $m = 3$, $k_c = 10$

Figure 5. Normalized Flooding results for DAPA model.

When a node of degree k_{min} receives a message, the node forwards the message to all of its neighbors excluding the node forwarded the message in the previous step. When a node with larger degree receives the message, it forwards the message only to randomly chosen k_{min} of its neighbors except the one which forwarded the message.

The NF search algorithm is based on the minimum degree in the network. The fixed minimum degree is equal to m by definition in PA and HAPA, whereas in CM an DAPA it is not guaranteed that the minimum degree will be m. In CM, deletion of self-loops and multiple links reduce the minimum degree to values less than m down to 1. In DAPA, however, the minimum degree might be less than m because of the short range of horizon for some peers which are geographically far from other peers. But still since the ratio of nodes with degree less than m is small we ignored them and ran NF algorithm based on the predefined minimum degree value m.

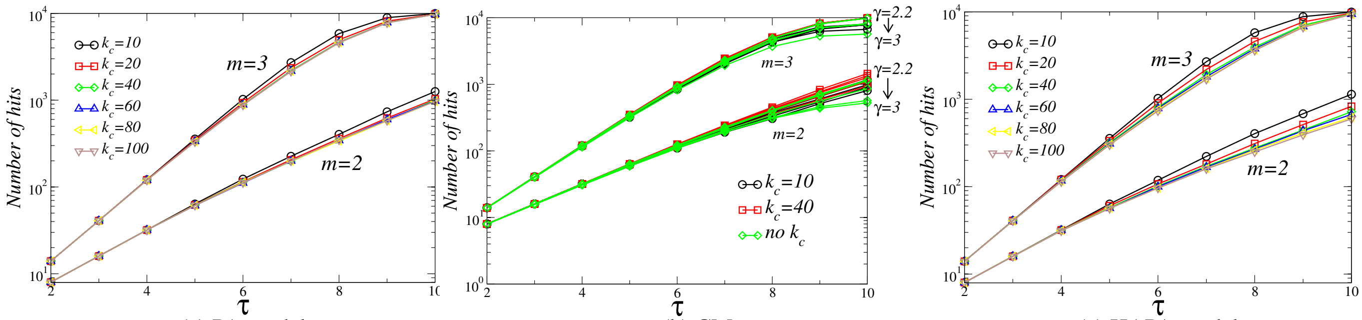
Random Walk (RW): RW or multiple RWs have been used as an alternative search algorithm to achieve even better granularity than NF. In RW, the message from the source node is sent to a randomly chosen neighbor. Then, this random neighbor takes the message and sends it to one of its random neighbors excluding the node from which it got the message. This continues until the destination node is reached or the total number of hops is equal to τ. RW can also be seen as a special case of FL where only one neighbor is forwarded the search query, providing the other extreme situation of the tradeoff between delivery time and messaging complexity. RW search is inherently serial (sequential), which causes a large increase in the delivery time [19, 13].

In particular, computer simulations performed on a generalized scale-free network with degree exponent γ=2.1, which is equal to the value observed in P2P networks, yield the result [4]: $T_N = N^{0.79}$.

5.2. Results

We simulated the three search algorithms FL, NF, and RW on the topologies generated by the four methods PA, CM, HAPA, and DAPA, and provide results all the combinations with various hard cutoffs. Through the PA, CM, HAPA, and DAPA methods, we generated topologies with 10000 nodes. We used cutoff values of 10 and 40 (or 50 in some cases), in addition to the natural cutoff, i.e., no hard cutoff. When generating DAPA topologies, we used τ_{sub} values of 2, 4, 6, 8, 10, 20, and 50 with expectation that larger τ_{sub} should yield better search efficiency. Minimum degree values (or m) in our topologies were 1, 2, or 3. We varied the τ values of search queries in FL up to the point we reach the system size and for NF/RW up to 10. To compare search efficiencies of RW and NF fairly in our simulations, we equated τ of RW searches to the number of messages incurred by the NF searches in the same scenario. Thus, for the search efficiency graphs of RW [e.g., Fig. 6] when τ is equal to a particular value such as 4, this means that the number of hits data-point corresponding to that $\tau = 4$ value is obtained by simulating a RW search with τ equal to the number of messages that were caused by an NF search using a τ value of 4. A similar normalization was done in [14].

In our simulation experiments, we observed that when there is no hard cutoff in the topology, the FL algorithm can achieve higher search efficiency by capturing more of the peers in the network for a specific τ value. Also, the effect of imposing a hard cutoff reduces when minimum degree in the topology is higher. We also observed that, for small values of cutoff, PA and HAPA give similar performances in FL, whereas for higher values of cutoff HAPA has better search results due to the star-like topology. The FL in DAPA is less efficient than in PA, although for higher values of τ_{sub} it gets closer to PA and efficiency of FL increases.



(a) PA model (b) CM (c) HAPA model

Figure 6. Random Walk results for PA, CM, and HAPA models for $m = 2$ and $m = 3$.

A minimum of three links for all peers eliminates negative effects of hard cutoffs: An interesting observation is that negative effect of hard cutoffs on the FL performance on the PA and HAPA topologies can be easily reduced to negligible values by increasing the number of stubs m (or connectedness). The number of stubs as small as 3 leaves virtually no difference between the search performances of overlay topologies with or without hard cutoffs. This result provides the guideline that *to achieve a better FL performance a requirement of having at least three links to the rest of the network will be adequate to assure that no one else in the network will need to maintain unbearably large number of links*. However, the necessity of complete or partial *global* information about the overall when constructing a PA topology is a major discouragement of using the PA and HAPA methods for overlay topologies of unstructured P2P networks.

There exists an interplay between connectedness and the degree distribution exponent for a fixed cutoff: As the DAPA method is a purely local method, it is more interesting to observe search performance on the DAPA topologies. We observed that when there is weak connectedness (e.g., $m = 1$), imposing hard cutoffs improves the search performance. This is due to the fact that hard cutoffs increase the connectedness of the topology by moving the links that were normally go to a hub in a topology without a hard cutoff. However, when the number of stubs is larger, we observe an interplay between the degree distribution exponent and connectedness for a fixed cutoff. We observe that improvement caused by hard cutoffs depend on the value of the hard cutoff, suggesting that reducing hard cutoff value hurts the search performance after a while. That is, *potential improvements by having smaller hard cutoffs diminishes as the performance starts to become dominated by the degree distribution exponent rather than the connectedness*. Another observation to be made is that impact of local information plays a major role in the search performance.

Hard cutoffs may improve search efficiency in NF and RW: More interestingly, for NF and RW, improvements due to having hard cutoffs are apparent in all three topology generation methods, including the PA topologies, regardless of the number of stubs m. This means that practical search algorithms like NF and multiple RWs are affected better by having hard cutoffs on the overlay topology. For NF, this is evidenced by Figs. 4(a) and 4(c) for the PA and HAPA topologies respectively. As it can be seen, having a little more local connectivity to the network by having a minimum of 2-3 links in every peer, the search performance increases rapidly for the same τ values. For RW, a very similar behavior is exhibited in Figs. 6(a) and 6(c), with only difference that effect of hard cutoffs is more apparent due to the fact that NF does better averaging of search possibilities. The observed behavior of RW illustrates how bad the effect of hard cutoffs can be on the search performance. It is intuitive that multiple RWs would perform more similar to NF in terms of performance.

More global information is more important when target connectedness is high: Figs. 5 shows the search performance of NF on DAPA topologies on semi-logarithmic scale when $m = 2$ and $m = 3$. We observe, again, that as the hard cutoff is getting smaller, the search efficiency improves regardless of the connectedness m. Also, having a little better connectedness (e.g. by comparing Fig. 5(a) against Fig. 5(b)) improves the search performance greatly. An interesting observation is that, when constructing the overlay topology, having more information (i.e. larger τ_{sub}) about the global topology (thus more scale-freeness in the overall topology) yields more important improvements on the search performance for topologies with more connectedness, i.e. larger m. This means that, for the purpose of constructing topologies with better search performance, when the target connectedness value is high one needs to be more patient and obtain as much information as possible before finalizing its links to the rest of the peers.

DAPA and HAPA models perform almost as optimal as the CM: An interesting characteristic to observe is how close the performances of DAPA and HAPA are to the best possible correspondent CM for the NF and RW search algorithms. Unlike the other topology construction mechanisms studied in this paper, CM achieves a perfect scale-freeness for a given target hard cutoff value, with the cost of global information. Specifically, topologies generated by the CM do not have big jumps at the hard cutoff values [e.g., Fig. 1(b)] in their degree distributions, in such a way that the links are configured in the perfect manner to assure that no node has links more than the target hard cutoff and the degrees of nodes follow exactly a power-law. This can be seen by comparing Fig. 2(a) with its counterparts Figs. 1, 2(b), and 3. As can be seen from Figs. 4(b) and 4(c), with connectedness $m = 2$ or $m = 3$, HAPA performs slightly worse than CM when using NF. Similarly, DAPA performance for moderate τ_{sub} values (e.g. 6) is very close to the optimal possible by the CM.

6. Summary and Discussions

We studied effects of the hard cutoffs peers impose on the number of entries they store on the search efficiency. Specifically, we showed that the exponent of the degree distribution reduces as hard cutoffs imposed by peers become smaller. We introduced new scale-free topology generation mechanisms (e.g., HAPA and DAPA) that use completely or partially local information unlike traditional scale-free topology generation mechanisms (i.e., PA and CM) using global topology information. We showed that topologies generated by our mechanisms allow better search efficiency in practical search algorithms like normalized FL and RW. Our study also revealed that interplay between the degree distribution exponent with a fixed hard cutoff and connectedness is likely to occur when using our mechanisms. We also showed that this interplay can be exploited by enforcing simple join rules to peers such as requiring each peer to have a minimum of 2-3 links to the rest of the unstructured P2P network.

Future work will include study of join/leave scenarios for the overlay topologies while attempting to maintain the scale-freeness of the overall topology. The challenge is to achieve minimal messaging overhead for join and leave operations of peers while keeping the scale-freeness in a topology with a hard cutoff.

Acknowledgment

This work is supported by the U.S. Department of Energy under Contract No. DE-AC52-06NA25396. M.Y. acknowledges support of the National Science Foundation with award number 0627039. We thank G. Korniss for a careful reading of the manuscript.

References

[1] Gnutella home page. http://www.gnutella.wego.com.
[2] The Freenet Project. http://freenetproject.org.
[3] The Napster home page. http://www.napster.com.
[4] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman. Search in power-law networks. *Physical Review E*, 64:046135, 2001.
[5] W. Aiello, F. Chung, and L. Lu. A random graph model for power law graphs. *Experimental Mathematics*, 10:53, 2001.
[6] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.
[7] E. Bender and E. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Comb. Theory A*, 24:296, 1978.
[8] M. Boguna, R. Pastor-Satorras, and A. Vespignani. Cut-offs and finite size effects in scale free networks. *The European Physical Journal B*, 38:205, 2004.
[9] M. Catanzaro, M. Boguna, and R. Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Phys. Rev. E*, 71:027103, 2005.
[10] E. Cohen, A. Fiat, and H. Kaplan. Associative search in peer-to-peer networks: Harnessing latent semantics. In *Proc. of IEEE INFOCOM*, 2003.
[11] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90:058701, 2003.
[12] S. Dorogovtsev and J. Mendes. Evolution of networks. *Adv. Phys.*, 51:1079, 2002.
[13] C. Gkantsidis, M. Mihail, and A. Saberi. Random walks in peer-to-peer networks. In *Proc. of IEEE INFOCOM*, 2003.
[14] C. Gkantsidis, M. Mihail, and A. Saberi. Hybrid search schemes for unstructured peer-to-peer networks. In *Proc. of IEEE INFOCOM*, 2005.
[15] K. Hui, J. Lui, and D. Yau. Small-world overlay p2p networks: Construction, management and handling of dynamic flash crowds. *Computer Networks*, 50(15):2727–2746, 2006.
[16] A. Iamnitchi, M. Ripeanu, and I. Foster. Small-world file-sharing communities. In *Proc. of IEEE INFOCOM*, 2004.
[17] J. Kleinberg. Navigation in a small world. *Nature*, page 845, 2000.
[18] A. Kumar, J. Xu, and E. W. Zegura. Efficient and scalable query routing for unstructured peer-to-peer networks. In *Proc. of IEEE INFOCOM*, 2005.
[19] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication strategies in unstructured peer-to-peer networks. In *Proc. of ACM Int. Conf. on Supercomputing*, 2002.
[20] S. Merugu, S. Srinivasan, and E. Zegura. Adding structure to unstructured peer-to-peer networks: The use of small-world graphs. *J. Parallel and Distributed Computing*, pages 142–153, 2005.
[21] M. Molloy and B. Reed. A critical-point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6:161, 1995.
[22] S. Ratnasamy, S. Shenker, and I. Stoica. Routing algorithms for DHTs: Some open questions. In *Proc. of IPTPS*, 2001.
[23] Q. Sun and H. Garcia-Molina. Partial lookup serrvices. In *Proc. of IEEE ICDCS*, 2003.
[24] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440, 1998.
[25] J. Xu. On the fundamental tradeoffs between routing table size and network diameter in peer-to-peer networks. In *Proc. of IEEE INFOCOM*, 2003.
[26] R. Zhang and Y. C. Hu. Assisted peer-to-peer search with partial indexing. In *Proc. of IEEE INFOCOM*, 2005.